American Journal of Epidemiology
Vol. 140, No. 9

Printed in U.S.A.

Is There Is or Is There Ain't No Baby?[1]: Dr. Shapiro Replies to Drs. Petitti and Greenland

Samuel Shapiro

I interpret Greenland's (1) assertion that I wish to ban meta-analysis as poetic license, and I thank him and Petitti (2) for their thought-provoking responses. To me, one of the most remarkable features of this symposium is that there is little disagreement among us concerning the synthetic meta-analysis (to use Greenland's expression) of nonexperimental data. Greenland and Petitti have not challenged any of the examples that I criticized (3), nor have they offered other examples that they consider valid. What that implies, by extension, is that we agree that the great bulk of the meta-analyses published to date are of questionable validity.

The large area of agreement considerably simplifies consideration of our disagreements. The responses raise three issues that are at the crux of the argument. First, Greenland makes a valuable distinction between synthetic meta-analyses whose purpose is to produce a single summary risk estimate, and those "aimed at testing criticisms of study results and identifying patterns or trends in [those] results." He considers synthetic meta-analysis to be misleading, except in those rare instances in which all studies are in agreement (and in that case, meta-analysis is superfluous). He even proposes that when study results are so heterogeneous as to require specification of a random-effects term, they are meaningless. On that matter, there is not a hairsbreadth of difference between us.

By contrast, Greenland considers the quantitative identification and description of how and why studies agree and disagree with each other to be the valid and worthwhile potential contribution to knowledge offered by meta-analysis. That is, if I interpret him correctly, meta-analysis can be used for the purpose of quantitative literature review. I have my doubts about that application as well, but if it were the only one, it is unlikely that we would be engaged in this symposium: by far the majority of the published meta-analyses have been of the synthetic type. Nevertheless, it is worth considering why we disagree with regard to meta-analysis as literature review. Because of the unquantifiable but important subtleties involved in any such review, I question whether quantitative methods can ever be as thoroughgoing, probing, and informative as qualitative methods.

The relative merits of the two approaches are well illustrated by the coffee–myocardial infarction controversy (4) to which Greenland refers to support his argument. His meta-analysis shows that the early case-control studies and cohort studies disagreed, with the consequence that the association was imputed to case-control biases; but that conclusion proved to be premature, because later cohort studies documented positive associations.

My reasoning, based on a qualitative review of the evidence, is as follows: if coffee increases the risk of infarction, the biologic mechanism might well be the acute adverse effects of caffeine on cardiovascular function (5–7); based on pharmacodynamic

Received for publication May 20, 1994, and in final form June 10, 1994.

[1] With the author's apologies to the late Louis Armstrong.

From the Slone Epidemiology Unit, 1371 Beacon Street, 3rd floor, Brookline, MA 02146. (Reprint requests to Dr. Shapiro at this address.)

Vol. 140, No. 9
ıd Public Health Printed in U.S.A.

'o Replies

natter, there is not a hairs-
ence between us.
Greenland considers the
tification and description
studies agree and disagree
o be the valid and worth-
ontribution to knowledge
nalysis. That is, if I inter-
ly, meta-analysis can be
ose of quantitative litera-
y doubts about that
ll, but if it were the only
that we would be engaged
m: by far the majority of
eta-analyses have been of
. Nevertheless, it is worth
we disagree with regard to
literature review. Because
able but important subtle-
ny such review, I question
tive methods can ever be
, probing, and informative
thods.
erits of the two approaches
ed by the coffee–myocar-
ontroversy (4) to which
to support his argument.
is shows that the early
ies and cohort studies dis-
consequence that the asso-
ted to case-control biases;
on proved to be premature,
hort studies documented
ions.
based on a qualitative re-
nce is as follows: if coffee
farction, the biologic
t well be the acute adverse
ie on cardiovascular func-
ed on pharmacodynamic

considerations, it is likely that such effects would be transitory, subsiding soon after cessation of intake, and dose related. Under those hypotheses, the case-control studies that examined consumption soon before the event and had sufficient power to assess heavy intake were more likely to be valid than the early cohort studies that examined intake years before the infarcts occurred, and that had insufficient power to assess heavy intake. The later cohort studies overcame those limitations. The valid studies, case-control and cohort, documented dose-related associations between coffee intake and myocardial infarction. The convergence of methodologically varied studies on the same relatively invariant association favors, but does not establish, causality.

I argue that my conclusions, derived from a qualitative assessment of the individual studies, including the nonepidemiologic literature, are more informative than Greenland's conclusions derived from a meta-analysis. But that is a matter of opinion. Perhaps all that the different approaches really reveal is that there is more than one way to skin a cat, and that it may be more useful to take the view that the quantitative and qualitative perspectives complement each other. In any event, however, the real concern is with the synthetic uses of meta-analysis, and on that topic we agree. In the remainder of this response, the term *meta-analysis* will be used as shorthand for synthetic meta-analysis.

The second issue raised in Greenland's and Petitti's responses is the argument that while individual nonexperimental studies, especially those documenting small effects, have intrinsic limitations of bias and confounding, we nevertheless continue to perform them. If we reject meta-analysis because of the same limitations, it is illogical not to reject the individual studies as well. That is, if we reject meta-analysis, we must by the force of that logic reject the whole corpus of nonexperimental epidemiology, at least as it applies to small effects.

That reasoning misses the point. We continue to perform individual studies in the hope that less equivocal results will follow from fresh approaches to a problem: for example, better definition of the hypothesis, making it more testable; or reduction of misclassification; or evaluation of possible dose/response effects over a greater range than in previous studies; or evaluation of the hypothesis in a population in which there are fewer competing risks; and so on. Meta-analysis offers no such opportunities, and it cannot bring any fresh or original insights to bear on a problem.

To be sure, sometimes we get stuck, and repeated studies all produce equivocal, low-magnitude effects. That is exactly the situation in which meta-analysis is most tempting. In some instances, there may even be strong *a priori* grounds for believing that an effect, although small, may be real, and if real, of great public health importance (e.g., cancer risk in relation to water chlorination (8)). However tantalizing that possibility may be, it is not a justification for meta-analysis; rather, if the suspicion is well grounded, it is a justification for further research that improves on the work performed to date.

Again, Greenland provides a useful illustration of the issues by referring to one of our studies of alcohol and breast cancer (9). That example is especially instructive because, in an earlier report, we had identified a weak association (10), and under a causal hypothesis, there were clear ways in which subsequent studies could be designed more rigorously. For example, the association arose in the course of performing multiple comparisons and may have been a statistical fluke; if so, a repeat study should have been null. Alcohol consumption was poorly defined in terms of amount and timing; assuming causality, better definitions should have yielded higher risk estimates.

For these reasons, we performed an additional independent study that made good on some of the defects (9). The result was null. That finding, together with a qualitative review of the literature (11), leaves me skeptical as to any possible causal relation. On the other hand, the two meta-analyses that I referred to (12, 13) reached the opposite conclusion. Readers interested in

reaching their own judgment as to which of the two approaches is the more valid might wish to compare the meta-analyses with the review.

The third issue raised by the responses is the *post hoc, ergo propter hoc* argument that since the new fashion of meta-analysis is unlikely to wither on the vine, we had better make the best of it. Meta-analysis will become an established part of the academic curriculum; there will be a cornucopia of funding for grants; and government departments will continue to make public health decisions, often misguided ones, based on the results of meta-analyses. We can play a role in keeping everything within reasonable bounds. Petitti also argues that qualitative review is an academic backwater, whereas meta-analysis has a certain appeal, and that we should do whatever we can to ensure that the limitations are recognized and the techniques used responsibly. I disagree: Bad science, however politically correct it may be for the moment, should be discredited as bad science.

The main issues having been responded to, a few remaining matters also call for response. First, Greenland offers elegant arguments that the definition of what constitutes a weak effect should vary according to context; and of course he is right in arguing that a relative risk of well below 2.0, or even 1.5, for a low exposure level can sometimes be interpreted as causal if higher levels of exposure produce higher risks (as with smoking). Nevertheless, in situations in which the highest risk that can be identified for any exposure category is less than twofold, I think my definition, although less elegant, is not materially different from his.

Second, Greenland offers some comments about data pooling, that is, the aggregation of "raw data" (as opposed to the meta-analysis of published data) from a range of studies. I did not consider data pooling because it fell outside the limits specified for the symposium. I am glad that the matter has been raised, however, because Greenland points out some important defects in data pooling, and also because there are a great many more that still need to be pointed out. The fashion has been accepted uncritically, and another symposium is needed.

Third, Petitti refers briefly to the application of meta-analysis to randomized controlled trials. Again, that topic was outside the scope of the symposium, and it deserves extensive debate in its own right.

Fourth, Petitti agrees with my criticisms of published meta-analyses, but she nevertheless feels that I am throwing out the baby with the bathwater. She argues that a "judgment about [the] promise [of meta-analysis] should not be based on the early studies that used the method." If so, where are the later studies that fulfill that promise? On conceptual grounds, can we expect there will be such studies?

Finally, I believe we are now confronted by a major educational dilemma. In recent years, meta-analysis has been uncritically embraced by many as a panacea. The scale of that embrace is unprecedented. There is hardly a medical journal in which it has not been claimed at some point that a "meta-analysis has shown that A causes B," or words to that effect. It took years to gain acceptance for the idea that $p < 0.05$ does not by itself indicate causation; it will probably take at least as long to drive home the limited interpretability, if not the lack of interpretability, of meta-analysis. One reason for the difficulty is that we have not yet set our own house in order: there are still too many epidemiologists who are willing to equate data aggregation with truth.

REFERENCES

1. Greenland S. Can meta-analysis be salvaged? Am J Epidemiol 1994;140:783–7.
2. Petitti DB. Of babies and bathwater. Am J Epidemiol 1994;140:779–82.
3. Shapiro S. Meta-analysis/shmeta-analysis. Am J Epidemiol 1994;140:771–8.
4. Greenland S. A meta-analysis of coffee, myocardial infarction and coronary health. Epidemiology 1993;4:366–74.
5. Robertson D, Frolich JC, Carr RK, et al. Effects of caffeine on plasma renin activity, catecholamines and blood pressure. N Engl J Med 1978;298:181–6.

6. Dobmeyer DJ, Stine RA, Leier CV, et al. The arrhythmogenic effects of caffeine in human beings. N Engl J Med 1983;308:814–15.
7. Robertson D, Curatolo PW. The cardiovascular effects of caffeine. In: Dews PB, ed. Caffeine: perspectives from recent research. New York, NY: Springer-Verlag, 1984.
8. Morris RD, Audet A-M, Angelillo IF, et al. Chlorination, chlorination by-products and cancer: a meta-analysis. Am J Public Health 1992;82:955–63.
9. Rosenberg L, Palmer JR, Miller DR, et al. A case-control study of alcoholic beverage consumption and breast cancer. Am J Epidemiol 1990;131:6–14.
10. Rosenberg L, Slone D, Shapiro S, et al. Breast cancer and alcoholic-beverage consumption. Lancet 1982;1:267–70.
11. Rosenberg L, Metzger LS, Palmer JR. Alcohol consumption and risk of breast cancer: a review of the epidemiologic evidence. Epidemiol Rev 1993;15:133–44.
12. Longnecker MP, Berlin JA, Orza MJ, et al. A meta-analysis of alcohol consumption in relation to risk of breast cancer. JAMA 1988;260: 652–6.
13. Longnecker MP. Alcoholic beverage consumption in relation to risk of breast cancer: meta-analysis and review. Cancer Causes Control 1994;5:73–82.